Survival Curve Convergences and Crossings:
How frequent are they in medical research? A Study of five medical journals

Abreham Mengesha Zewdu

Institute of Health Management and Health Economics

Faculty of Medicine

UNIVERSITETET I OSLO

October, 2009

Preface

First and for most I thank Jesus Christ for giving me the strength to bear all the difficulties in the course of producing this research paper.

My special thanks go to my primary advisor Professor Ivar Kristiansen who encouraged and guided me at all stages of this project. Without his support and encouragement it would have been difficult to realise it. Professor Ivar is the most honest and straightforward person I ever worked with. He does not hesitate the slightest to speak his mind and enthusiastic to listen and to learn from his own student. His critical comments and suggestions throughout the process are of immense importance for the accomplishment of this paper.

My grateful thanks go to my co-adviser Henrik Støvring (PhD) for his valuable comments and suggestions especially at the initial stages of this project. I also thank Tron Anders Moger (PhD) for his valuable statistical advice.

My deepest tanks go to my lovely wife Liah Amanuel and to my kids Ruth and Naomi for their love, understanding and unwavering support throughout the process.

My last but by no means least thanks go to my colleges and very good friends Donata and Haile for their constructive comments through out the project process.

Abreham Mengesha Zewdu
Oslo, September, 2009

Abstract

Background: In medical publications, effectiveness of health interventions for chronic diseases is usually expressed as absolute risk reduction (ARR), number-needed-to-treat (NNT) or relative risk reduction (RRR). These measures are estimated at one point in time and require the hazard rates to be constant over time in order to yield information that is representative for the whole interventions period. Measurement at one point in time may not be adequate if the relative hazard for the event of interest (typically death) varies with time. Individual patient data are required to estimate hazard rates and relative hazards. However, survival curves may be used to make inferences about relative hazards. Crossings and/or convergences of survival curves after they have diverged clearly indicates the relative hazard is not constant.

Objectives: To explore how frequent survival curves do converge and/or cross in medical research and to investigate determinants of convergences and crossings.

Design: Review of all publications that included survival graph during 2007 in five major peer-reviewed medical journals. The following data were extracted: type of disease, type of exposure, number of comparator groups, number of pairwise comparisons, type of primary and secondary end-points, sample size, maximum follow-up time, survival curve convergences, survival curve crossings, type of epidemiologic study design, result of log-rank test (if reported), and country in which the study was undertaken.

Sample: 177 publications from Annals of Internal Medicine (AIM), British Medical Journal (BMJ), Journal of the American Medical Association (JAMA), New England Journal of Medicine (NEJM) and The Lancet.

Results: 78% of the publications had survival curve convergences and 42% survival curve crossings. The proportion of survival curve convergences and crossings varied across disease type, intervention type, number of comparator groups, number of pairwise comparisons, types of primary and secondary endpoints, sample size, study design and length of follow-up time. In multivariate logistic regression, survival curve convergence was positively associated with ‘more than one pairwise comparison’ (OR 3.7, 95% CI 1.3-10.8) and ‘death as a secondary endpoint’ (OR 8.1, 95% CI 1.1-65.5). No association was found between survival curve crossings and any of the explanatory variables.

Conclusion: Survival curve convergences and crossings are common phenomena in medical research. The phenomena seem not to be associated with particular study characteristics. The results warrant care in making inferences about the effectiveness of interventions for chronic diseases on the basis of measurement at a single point in time.

Table of content

List of Tables

List of figures

Abbreviations

AIM	Annals of Internal Medicine
ARR	Absolute risk redaction
BMJ	British Medical Journal
CBA	Cost benefit analysis
CEA	Cost effectiveness analysis
CI	Confidence interval
CUA	Cost utility analysis
CVD	Cardiovascular disease
JAMA	The Journal of the American Medical Association
NCD	Non communicable diseases
NEJM	The New England Journal of Medicine
NNT	Number-needed-to-treat
OR	Odds ratio
QALY	Quality adjusted life year
RCT	Randomised controlled trial
RR	Relative risk
RRR	Relative risk reduction
WHO	World Health Organization
WTP	Willingness to pay

1. Background

Non-communicable diseases (NCDs) are the leading causes of death and disability globally. In 2002, for example, the four major diseases; cardiovascular diseases (CVD), cancer, diabetes, and chronic respiratory illnesses together accounted for 29 million deaths worldwide (Yach et al, 2004). The global burden of NCDs is expected to increase in the future partly due to the decline in mortality from communicable diseases and partly due to the increase in the proportion of the aging population in middle and low-income countries (Alwan et al, 2001). In Europe, NCDs accounted for 86% of all deaths. Cardiovascular disease (52%), cancer (19%), chronic respiratory disease (4%), diabetes (2%), and other chronic diseases (9%) were the major causes of deaths (WHO, 2005).

In order to tackle the problem, countries have taken a number of preventive and curative steps in the form of adoption of multilateral strategies (where WHO plays a leading role), national policy directives, and clinical and community levels of interventions. Equally important steps have been to measure the benefits of interventions (intervention effectiveness). This is important for at least two reasons. First, in a clinical setting, measures of intervention effectiveness provide clinicians with useful information on whether or not a given intervention is effective. Second, in managerial settings, measures of intervention effectiveness are important to ascertain whether or not a given intervention is cost-effective. There are various ways of measuring health interventions effectiveness. One widely used measure is gain in survival time (life years) which is usually estimated from survival curves. Researchers make use of survival curves to estimate the life year gain of a new treatment compared to its comparator (either a placebo or a conventional treatment). They also use survival curves to predict how long time members of a given intervention group are expected to survive. Moreover, survival curves are an integral part of cost-effectiveness analyses that are used for resource allocation and priority setting in health care.

What actually prompted this very study is that, in medical publications effectiveness of health interventions for chronic diseases are usually expressed as absolute risk reduction (ARR), number-needed-to-treat (NNT) or relative risk reduction (RRR). These measures are estimated at one point in time and require the hazard rates to be constant over time in order to yield information that is representative for the whole interventions period. The potential danger of such estimate is the tendency to aggregate the effect of treatment as an average

benefit, disregarding the possible variation over time. Interventions effectiveness measures based on a single point in time may be inadequate if the relative hazard for diverse event (typically death) varies with time. While it requires individual patients' data to judge how much the relative hazard varies over time, such variation can be inferred to exist from the survival curves that converge and/or cross. The purpose of this study was to investigate how often survival curves do converge and/or cross and explore the main determinants. The following section deals with concept of survival analysis followed by the objectives of the study. The third section presents the methods and material used in the study. The fourth section presents the results of the study, and the final section deals with discussion and conclusions.

2. Theory and Concept

2.1 What is survival analysis?

Survival analysis is a statistical method for analysing survival data that arise frequently in the medical setting from both clinical randomized controlled trail (RCT) and epidemiological studies. Survival analysis is appropriate when the interest is, for example, to study the time duration from a well-defined origin, such as initiation of a treatment or the diagnosis of a disease, until the incidence of some particular event of interest, such as death or recovery. Although there is usually a clearly defined endpoint of interest, the origin may not be well-defined. For example, if it simply refers to time of entry, which may be a rather arbitrary feature of the study design, *i.e.* an extended period of patient recruitment (Figure 1). However, it is important to realize that although patients may be recruited at different times, the survival curve should represent the experience of each patient from the time of exposure (intervention, treatment, risk factor etc). The survival technique enables researchers to follow subjects over time and observe the event of interest (as in A, C, and D in Figure 1). However, the event of interest may not be observed in all subjects. Some subjects may withdraw before the event of interest was observed (E) or the study was completed before the event of interest had been observed (B). In either case, a complete survival time is not available. Thus, survival data with such incomplete observation are said to be right censored. All that is known about a censored observation is partial information collected up to the point in which the subjects were part of the study, i.e. while we do not know the exact time the event occurs for the individual, we do know it exceeds the time of censoring (Altman, 1991, Parmar and Machin, 1996).

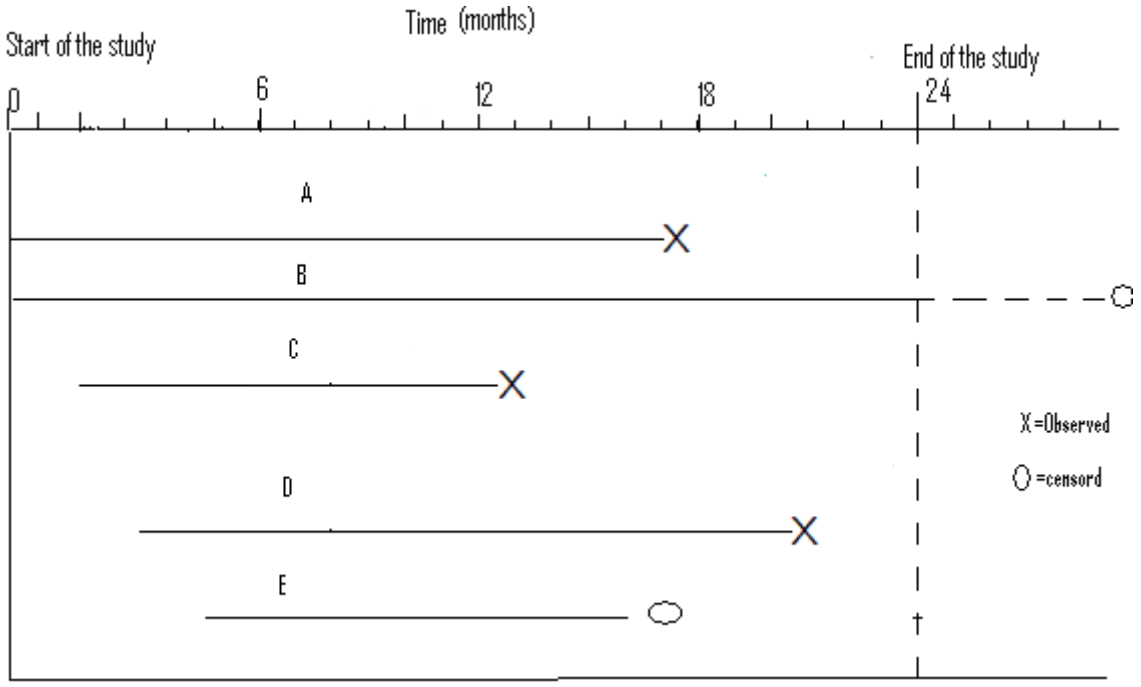

Figure 1. Different entry point and subject experience in a typical survival analysis: A, C, &D are fully observed; B & E are right censored. E due to drop-out before end of study. (Adapted with modification from Altman.1991).

Survival data collected in such a way (as depicted in Figure 1) are used to estimate survival probabilities, to compare survival experiences in different groups, and to assess the relationship between explanatory variables and survival time. Such analysis is done through the use of survival curves or life tables, as well as dedicated regression techniques.

2.2 What is a survival curve?

A survival curve depicts survival probability. The theoretical bases of a survival graph are survivor function *S(t)* and hazard function *h(t)* (Altman, 1991). Survivor function *S(t)* expresses the probability that an individual will survive up to and including time *t*.

$$S(t) = P(T>t)$$

Where, *t* is a given point in time, *T* is a random variable denoting the time of an event, and *P* is probability. Put it differently, the formula implies that the survival function is the probability that the time of an event such as death occurs later than some specified time. The hazard function *h(t)* expresses the probability that an individual experiences the event instantaneously after a time *t* given the subject has not yet experienced the event. It can

increase, decrease, or remain constant over time. A higher hazard rate is associated with lower survival (Kirkwood & Jonathan, 2003).

The survival curve starts at 100% where all subjects are alive and declines with each event until the last event of interest or outcome has been observed (Figure 2). When it is estimated with the Kaplan-Meier technique, it is a step function in which the survival probabilities remain unchanged up to the time of the first event, such as year 1, and followed by a step down for each event up until year 6, beyond which the survival probabilities were unchanged up to year 12 (this is because no event was observed from year 6 until year 12). Effectively, we can not measure any survival time beyond year 12, as this is the end of the trial. The area under the curve represents the average survival time.

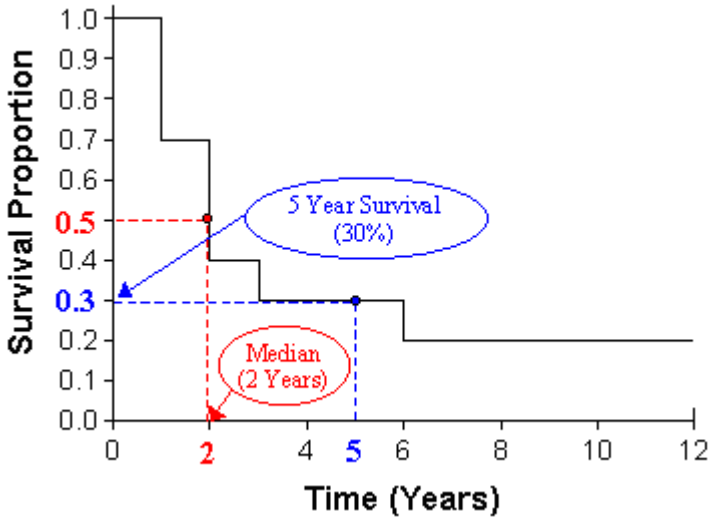

Figure 2. Graphic illustration of a survival curve. Adapted from Cancer guide homepage (http://www.cancerguide.org)

Depending on the data available survival curves can be estimated using either an actuarial method or the Kaplan-Meier method. In the actuarial method, the x-axis is divided into regular intervals, and percent survival is calculated for each interval, as is done in the life table method, on the assumption that individuals are lost to follow-up halfway through the interval. In the Kaplan-Meier method, the survival probabilities are recalculated each time a patient has an event, whereas censorings only remove individuals from the risk set used to calculate probabilities at subsequent event times (Wilson, 2004; Altman 1991; Kirkwood & Jonathan 2003; Parmer & Mahin, 1995).

2.3 Comparing two survival curves

One of the major purposes of survival analysis is to compare the survival experiences of two or more groups. In clinical trials, it is widely employed to compare the result of two or more interventions to decide which treatment is more effective. By studying two survival curves, researchers analyse whether the survival time is the same or different in two groups (Figure3). For example, the group under Treatment A enjoyed a higher median survival time (4.4 years) than the group under Treatment B (2.5 years). This means that Treatment A provided a gain of almost two life years in terms of median survival time. Similarly survival curves can be compared at any other percentile. It is also possible to conclude that patients under treatment B have worse survival than those under treatment A throughout the study period. This is evidenced from the pattern of the survival graph that the survival curve depicting Treatment A lies above the curve depicting Treatment B throughout the follow-up period. Studying the figure also roughly informs us whether or not the relative hazard varies over the course of the follow-up period. If two survival curves do converge or cross each other as in Figure 4, the relative hazard is not constant over time. Even if survival curves do not converge or cross each other, the relative hazard may not be constant, but if they converge or cross we know for sure that the relative hazard is not constant. Finally, it should be noticed that all survival curves by definition coincide at time zero and when all have experienced the event. When we refer to convergences and/crossings we, thus, referring to time point in-between, i.e. when some, but not all subjects experienced the event.

The question is, however, how to measure effectiveness of an intervention along a survival curve. Effectiveness of intervention particularly for chronic diseases can be measured horizontally, vertically, or through a combination (Kristiansen and Gyrd-Hansen, 2006, P 674). By horizontal measure, we mean that by drawing a horizontal line that connects the two survival curves (Figure 3), it is possible to capture the percentage of patients alive for some particular time period. Although the horizontal effectiveness measure is conventionally expressed as median survival (the survival of 50% of the population in each group) as in our

example above, it is possible to measure at any survival proportion of the population.

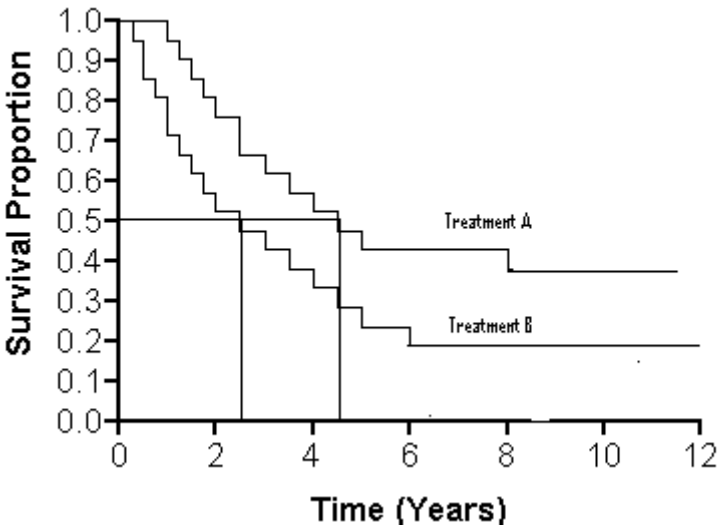

Figure 3. Comparison of two survival curves. Adapted from the Cancer Gide webpage (http://www.cancerguide.org)

For vertical effectiveness measures we draw a vertical line between two survival curves, and register the proportion of patients who have experienced the event of interest, such as death. And the relation between the survival in the two groups can be expressed in terms of relative risk (RR), absolute risk reduction (ARR), number-needed-to treat (NNT) or odds ratio (OR). These measures can be taken at any point along a survival curve though usually they are taken at the end of the trial. The third way of measuring the outcome of the intervention is by measuring the area between the two survival curves. This measure combines both the horizontal and the vertical measures, and intends to capture all the information on the difference between the two survival curves. It is generally referred as the average postponement of an adverse event.

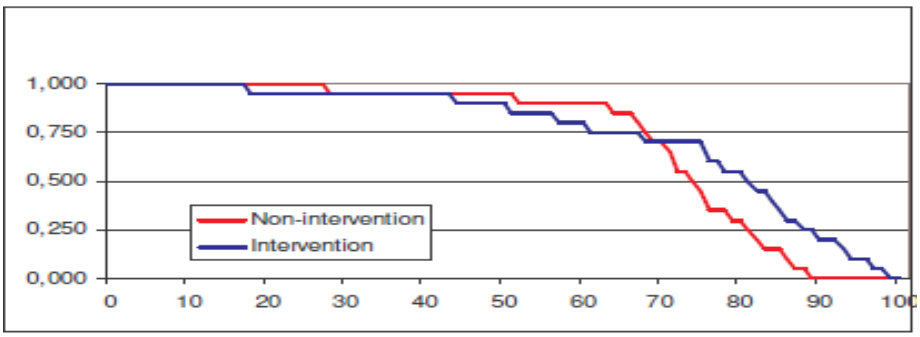

Figure 4 Illustration of survival curve convergences and crossings. Adapted from (Kristiansen and Gyrd-Hansen 2006).

However, comparison between groups may be inaccurate and difficult to interpret. Some of the reasons for this:

The difficulty in singling out whether or not the difference between two groups is due to treatment differences rather than patient differences; 2) different types of measures capture different aspects of the health effect; 3) Logrank tests are used when they are inappropriate, for example, when survival curve cross; and 4) the values of effect measure are highly dependent on when they are measured.

Most health effects are measured at the end of the trial or at some other single arbitrary time point. However, such a measure may not provide a relevant comparison of the total survival experience between groups, in particular, when it fails to capture: 1) the difference in survival across time, and 2) survival curves that may converge or crossover. The purpose of our study was, therefore, to investigate survival curve convergences and crossings in medical research. To the best of our knowledge no many studies have explored how often survival curves do converge or cross and its determinants. Our aim was to capture survival graphs from scientific publications presented in major medical journals in pre-specified period of time and investigate to what extent survival curves do converge and cross each other. Specifically, the objectives of our study were:

- To explore how frequent survival curves do converge and/or cross;
- To explore determinants of survival curves convergence and crossing.

In doing so, we hypothesised the followings:

- Convergences and crossings are rare phenomenon in medical research
- Convergence and crossing will be explained by disease type, exposure type, sample size, member of comparator groups, number of pairwise comparisons, type of study design, length of follow-up time, and type of endpoint.

3 Methods and Materials

We based our research on the following five major peer-reviewed medical journals: Annals of Internal Medicine (AIM), British Medical Journal (BMJ), The Journal of the American Medical Association (JAMA), The New England Journal of Medicine (NEJM), and The Lancet. For each journal we browsed through all 2007 issues and identified all publications that included survival graphs.

3.1 Inclusion and exclusion of publications

Among the publications with survival graphs, we excluded the following: 1) publications with a survival graph having no comparator group (survival graph with a single curve), and 2) publications with a survival graph that were reproduced from other journals without complete information on methods and results.

3.2 Study sample and Data extraction

We identified 177 publications which met our study criteria. In the process of data extraction we observed the following patterns:

1. Publications with no comparator group (a survival graph with a single curve) (Figure 5).

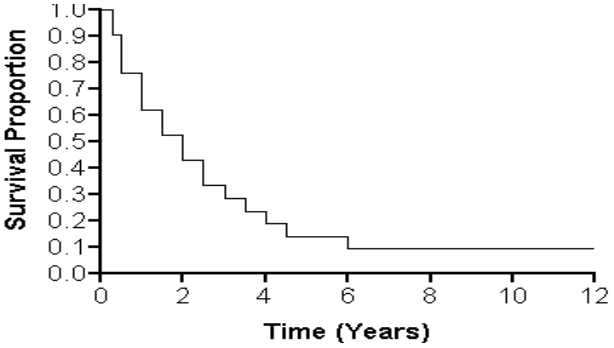

Figure 5. A survival graph with out a comparator group. Adapted from the Cancer Gide webpage (http://www.cancerguide.org)

2. Publications with a single endpoint, with two comparator groups presented by a single survival graph (Figure 6). For such types of survival graphs there is only a single possible pairwise comparison.

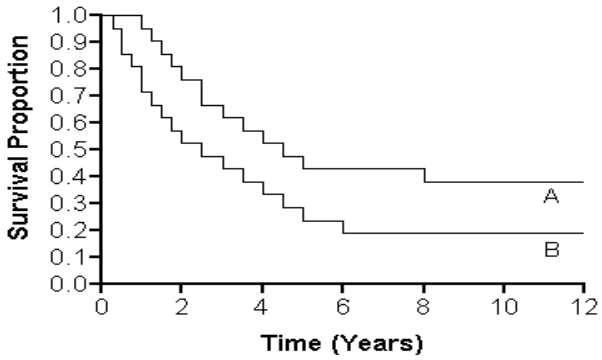

Figure 6. A survival graph with two comparator groups. Adapted from the Cancer Gide webpage (http://www.cancerguide.org)

3. Publications with a single endpoint but with more than two comparator groups presented by a single survival graph (Figure 7). For such types of survival graphs the potential pairwise comparison varies depending on the number of comparator groups (*i.e.* survival curves).

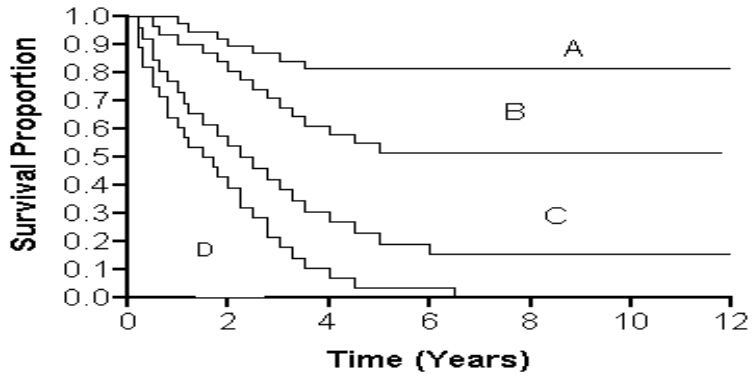

Figure 7. A survival graph with more than two comparative curves. Adapted from the Cancer Gide webpage (http://www.cancerguide.org)

In our example, we have potentially six pairwise comparisons [(A *vs.* B), (A *vs.* C), (A *vs.* D), (B *vs.* C), (B *vs.* D), (C *vs.* D)].

4. Publications with multiple endpoints and two comparator groups for each endpoint. Here, usually the number of pairwise comparison is equal to the number of endpoints (Figure 8).

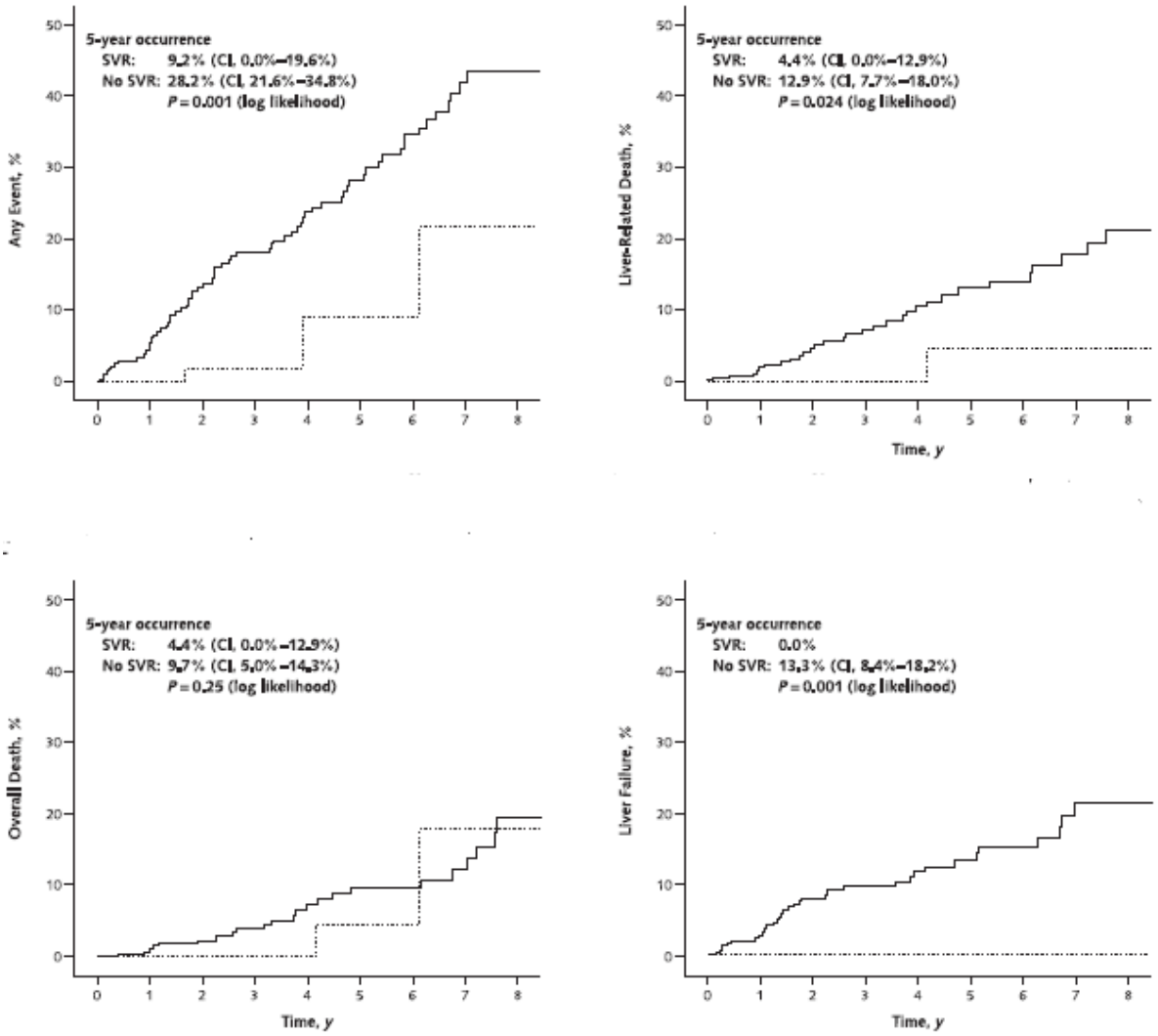

Figure 8. Multiple end points each represented by a single survival graph. Adapted from Veld et al (2007).

5. Publications with multiple endpoints and more than two comparator groups presented by multiple survival graphs (Figure 9). The potential number of pairwise comparison may be very large depending on the number of endpoint and the number comparator groups.

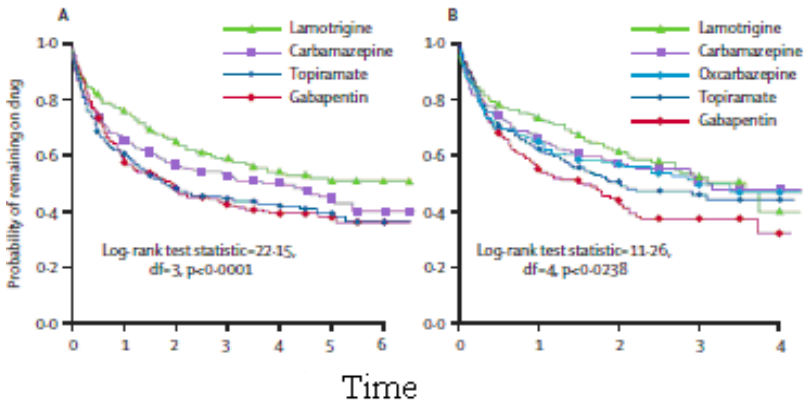

Figure 9. Multiple endpoint with more than two comparative curves presented by more than one survival graphs.Adapted from Marson et al (2007).

As some of the publications were included more than one survival graphs, the total number of survival graphs observed was greater than the number of publications included in the study. We used the followings rules when extracting the data:

- If death or a combination of death and other outcomes were a primary or secondary endpoint, data from this graph was recorded.
- In case both the primary endpoint was a non-death outcome or a combination of non-death outcome, we chose a graph representing a combined endpoint, if available; otherwise, every graph was recorded. With these criteria some publications were represented by one graph, but some were represented by two or more graphs. The following data were extracted from each of the 177 identified publications:
    - Type of disease,
    - Type of exposure,
    - Number of comparator groups,
    - Number of pairwise comparison,
    - Type of primary and secondary endpoints,
    - Sample size,
    - Maximum follow-up time,
    - Number of survival curve convergences,
    - Number of survival curve crossings,
    - Type of epidemiologic study design,
    - Result of log-rank test (if reported),
    - Country in which the study was undertaken.

Survival curve convergences and crossings data were extracted directly from survival curves while the remaining data were extracted from the manuscript of the publications.

3.3 Data Analysis

We broadly classified publications as shown in Table 2. The data were analyzed by means of contingency tables and simple descriptive statistic. Determinants of convergences and crossings were analyzed with logistic regression (Tables 8 & 9). The dependent variables convergence and crossing were dichotomized into convergence (=1)/no convergence (=0) and crossing (=1)/no crossing (=0) and defined as follows: for Survival curve convergence; (=0) represent none of the survival curves in a give publication were converged, and (=1) represent at least two survival curves were converged once. For survival curve crossings; (=0) represent none of the survival curves in a given publication were crossed while (=1) represent at least

two curves were crossed once. The following predictor variables were used and categorized as follows:

- Disease type: cancer (=0), cardiovascular diseases(=1), infections (=2) and all others (=3);
- Exposure type: pharmaceuticals (=0), surgery (=1), diagnostic (=2), procedures and devices (=3), and all others(=4);
- Number of comparator groups[1]: two comparator groups (=0) and more than two comparator groups (=1).
- Number of pairwise comparison[2]: one pairwise comparison (=0) and two or more pairwise comparisons (=1).
- Primary endpoint: death (=0) and non-death outcome (=1).
- Secondary endpoint: non-death (=0) and death from any cause (=1).
- Type of study design: randomized controlled trail (RCT) (=0), cohort study (=1) and all other (=2).
- Length of follow-up: 0-180 days (=0), 181-366 days (=1), 367-732 days (=2), 733-1830 (=3) and more than 1830 days (=4) of follow-up.
- Sample size: It was expressed per 100 units.

For independent variable all categories represented by (= 0) used as a reference. For all analyses, a 2-tailed P<.05 was considered statistically significant. Data were analyzed in SPSS version 16.0.

[1] Comparator groups: are two or more groups of patients randomized and exposed to different exposure types in a given trial.

[2] Pairwise comparison: is the number of comparison made between groups. Usually it is the function of the number of comparator groups in a given study and follows the following pattern;

Number of comparator	2	3	4	5	6
Potential number of pairwise comparison	1	2	6	10	15

4. Results

4.1 Descriptive statistics

We identified a total of 177 publications which met our inclusion criteria, 17 from AIM, 7 from BMJ, 44 from JAMA, 68 from NEJM, and 41 from The Lancet. The proportion of publications by geographic region varied from 5% for Africa to 39% for Multi-national (Table 1).

Table 1. Publications according to medical journal and Geographic region

	Geographic region					
Journal	America	Europe	Asia & Australia	Africa	Multi-national	Total (%)
AIM	7	5	1	0	4	17 (9.6)
BMJ	1	5	1	0	0	7 (4.0)
JAMA	22	7	1	2	12	44 (24.9)
NEMJ	13	9	6	2	38	68(38.4)
The LANCET	1	14	6	5	15	41 (23.2)
Total (%)	44 (24.9)	40 (22.6)	15 (8.5)	9 (5.0)	69 (39.0)	177 (100)

Variations in proportion of publications were observed across various study characteristics (Table 2). For disease type, the proportion of publications varied from 2% for pulmonary to 35% for cardiovascular. For type of intervention, the proportion of publications varied from 5% for 'risk factor' to 56% for pharmaceutical. For number of comparator groups, the proportion varied from 28% for 'more than two groups' to 72% for publications with 'only two groups.' For number of pairwise comparisons, the proportion of publications ranged from 39% for 'one pairwise' comparison to 61% for 'more than one pairwise' comparisons. For primary endpoint, the proportion of publications ranged from 2% for 'other' to 37% for 'single non-death outcome'. For secondary endpoint, the proportion ranged from 19% for 'death from any cause' to 81% for 'non-death outcome.' For study design, the range was from 11% for 'other' to 68 % for randomized controlled trial (RCT). For length of follow-up time, the proportion of publications ranged from 11% for '181-366' days of follow-up to 28% for 'more than 1830' days of follow-up. 50% of the publications reported Log-rank test.

Table 2. Publications according to study characteristics

Variables and Categories	N	%
Disease type	**177**	**100**
Cancer	47	27
Cardiovascular disease	62	35
Infection	20	11
Kidney disease	11	6
Pulmonary disease	4	2
Other	33	19
Exposure type	**177**	**100**
Pharmaceutical	99	56
Surgery	13	7
Diagnostic method	14	8
Procedure/Device	34	19
Risk Factor	8	5
Other	9	5
Number of comparator groups	**177**	**100**
Two Groups	128	72
>Two Groups	49	28
Number of pairwise comparison	**177**	**100**
One Pairwise	69	39
>One Pairwise	108	61
Primary endpoint	**177**	**100**
Death from any cause	56	32
Death + non-death outcome	34	19
Single Non-death Outcome	66	37
Combined Non-death Outcome	18	10
Other	3	2
Secondary endpoint	**177**	**100**
Non-Death Outcome	143	81
Death from any Cause	34	19
Study type	**177**	**100**
Randomized controlled trial (RCT)	121	68
Cohort Study	36	20
Other	20	11
Length of follow-up (in days)	**177**	**100**
0-180	34	19
181-366	19	11
367-732	20	11
733-1830	54	31
>1830	50	28
Log-rank value	**177**	**100**
Not Reported	89	50
Reported	88	50

Out of the total 177 publications, 138 (78%) had survival curve convergences and 74 (42%) had survival curve crossings. Variations in the proportion of survival curve convergences and crossings were observed across medical journals and geographic regions (Table 3). Survival curve convergences ranged from 68% for JAMA to 88% for AIM, and from 61% for America to 89% for Africa. The corresponding variations in proportion for survival curve crossings were 36% for JAMA to 57% for BMJ, and 37% for Europe to 56% for Africa.

Table 3. Survival curve convergences and crossings by types of journal and geographic region

		Survival Curve	
		Crossings	Convergences
Type of Journal	N	%	%
AIM	17	53	88
BMJ	7	57	71
JAMA	44	36	68
NEMJ	68	40	81
The LANCET	41	44	80
Total	177	42	78
Geographic Region	N	%	%
America	44	41	61
Europe	40	37	83
Asia & Australia	15	40	67
Africa	9	56	89
Multi-national	69	44	87
Total	177	42	78

Some variations in survival curve convergences and crossings were observed across disease type (Table 4). The lowest proportion of survival curve convergence was observed for pulmonary disease (75%) while the highest proportion was for kidney diseases (91%). For survival curve crossings, the lowest proportion was observed for kidney (27%) while the highest was for pulmonary (75%).

Table 4. Survival curve convergences and crossings according to disease type

		Crossings	Convergences
Disease Type	N	%	%
Cancer	47	38	77
Cardiovascular	62	39	77
Infection	20	55	80
Kidney Disease	11	27	91
Pulmonary Disease	4	75	75
Other	33	45	76
Total	177	42	78

Survival curve convergences and crossing by the types of exposure illustrated that the lowest proportion of convergence was observed for 'risk factor' (50%) while the highest proportion was observed for the intervention type 'other' (89%) (Table 5). For survival curve crossings the lowest proportion was observed for 'risk factor' (13%) while the highest proportion was observed for pharmaceuticals (48%).

Table 5 Survival curve convergences and crossings according to exposure type

		Crossings	Convergences
Exposure Type	N	%	%
Pharmaceutical	99	48	83
Surgery	13	39	85
Diagnostic	14	29	57
Procedure/Device	34	38	74
Risk factor	8	13	50
Other	9	44	89
Total	177	42	78

Survival curve convergences and crossings differed across types of primary endpoint (Table 6). The lowest proportion was observed for 'combined non-death primary outcome' (67%) while the highest proportion was for 'other' (100%). The corresponding lowest and highest proportions for survival curve crossing were 0% and 56% for 'other' and 'combined non-death outcome' respectively. For secondary endpoint, the proportion of convergences ranged from 73% for 'non-death outcomes' to 97% for 'death outcome'. However, no major variation was observed for crossing 38% for 'death outcome' to 43% for 'non-death outcome'.

Table 6. Survival curve convergences and crossings according to primary end-point

		Crossings	Convergences
Primary end-point	N	%	%
Death from any cause	56	36	75
Death + non-death outcome	34	47	94
Single non-death outcome	66	42	74
Combined non-death outcome	18	56	67
Other	3	0	100
Total	177	42	78

Survival curve convergences and crossings also differed across length of follow-up time (Table 7). The lowest proportion of survival curve convergences was observed for '181-366' days of follow-up (64%) while the highest was observed for '0-180' days of follow-up (85%). The corresponding figure for survival curve crossing was 26% and 53% for '180-366' and '0-180' days of follow-up respectively.

Table 7. Survival curve convergences and crossings according to length of follow-up time			
		Crossings	Convergences
Length of follow-up (in days)	N	%	%
0-180 (up to six months)	34	53	85
180-366 (up to a year)	19	26	64
367-732 (up to two years)	20	40	80
733-1830 (up to five years)	54	42	80
> 1830 (more than five years)	50	40	74
Total	177	42	78

Variations in survival curve convergences and crossings were observed across the number of comparator groups, the number of pairwise comparisons, and the type of epidemiological study design. The proportion of survival curve convergences was about the same for publications with 'more than two groups (76%)' and for publications with 'only two groups (79%).' Survival curve crossings ranged from 33% for publications with 'more than two groups' to 45% for publications with 'only two groups.' For the number of pairwise comparisons, survival curve convergences ranged from 71% for publications with 'only one pairwise comparison' to 82% for publications with 'two or more pairwise comparisons.' The proportions of survival curve crossings was almost the same for publications with 'only one pairwise comparison' (41%) to publications with 'more than one pairwise comparison (43%).' The proportion of convergences varied from 6% for the study design 'other' to 84% for randomized controlled trial (RCT). Crossings varied slightly from 35% for 'other' to 44% for randomized controlled trial (RCT).

4.2 Regression analysis

In bivariate analysis of survival curve convergences we found the following statistically significant predictors: 'diagnostic' *versus* 'pharmaceuticals exposure' (OR 0.28, 95% CI 0.09- 0.90), 'non-death' *versus* 'death' as a secondary endpoint (OR 11.9, 95% CI 1.58-90.37), 'cohort' *versus* 'RCT' study design (OR 0.37, 95% CI 0.16-0.87), 'others' *versus* 'RCT' study design (OR 0.28, 95% CI 0.10-0.78) (Table 8). However, no significant association was found between convergences and disease type, other categories of exposure type, number of comparators group, number of pairwise comparisons, type of primary endpoint, length of follow-up time, and sample size. No association was found between survival curve crossings and any of the predictor variables.

In multivariate analysis of survival curve convergences we found the following statistically significant predictors: ‘two or more pairwise’*versus*‘only one pairwise comparisons’ (OR 3.7, 95% CI 1.3-10.8), ‘non-death’ *versus* ‘death’ as a secondary outcome (OR 8.1, 95% CI 1.1-65.5) (Table 9). However, no statistically significant association was found between convergences and disease type, exposure type, number of comparators group, type of primary endpoint, length of follow-up time, study design, and sample size. No association was found between survival curve crossings and any of the explanatory variables.

Table 8. Bivariate Analysis of Survival Curve Crossings and Convergences

	95% CI for Odd-ratio (OR)							
	Crossings as dependent variable				**Convergences as dependent variable**			
Independent Variables	OR	Lower	Upper	*P*	OR	Lower	Upper	*P*
Disease Type								
Cancer (ref.)	1				1			
Cardiovascular	.99	.45	2.15	.95	.978	.40	2.38	.96
Infection	1.96	.68	5.67	.21	1.222	.34	1.43	.76
All others	1.30	.57	2.96	.53	1.290	.48	3.48	.62
Exposure Type								
Pharmaceuticals (ref.)	1				1			
Surgery	.69	.21	2.26	.54	1.140	.231	5.62	.87
Diagnostic	.44	.13	1.51	.19	.276*	.09	.90	.03
Procedure/device	.68	.31	1.52	.35	.576	.23	1.45	.24
All Others	.46	.15	1.41	.17	.498	.16	1.60	.24
Comparator Groups								
Two groups (ref.)	1				1			
Two or more groups	.58	.29	1.17	.12	.824	.38	1.79	.63
Pairwise comparison								
One pairwise (ref.)	1				1			
Two or more pairwise	1.08	.59	2.01	.79	1.912	.93	3.92	.08
Primary endpoint								
Death (ref.)	1				1			
Non-death Outcome	1.16	.64	2.12	.62	.602	.29	1.24	.18
Secondary endpoint								
Non-death (ref.)	1				1			
Death from any cause	.83	.39	1.79	.63	11.943*	1.58	90.37	.02
Study Type								
RTC (ref.)	1				1			
Cohort	.81	.38	1.75	.60	.373*	.16	.87	.02
Others	.96	.26	1.85	.46	.279*	.10	.78	.01
Follow-ups in Days								
0-180 (ref.)	1				1			
181-366	.31	.09	1.08	.06	.374	.10	1.45	.15
367-732	.59	.19	1.28	.36	.690	.16	2.93	.62
733-1830	.65	.28	1.56	.34	.674	.21	2.14	.50
>1830	.59	.25	1.42	.24	.491	.16	1.53	.22
Sample size								
Sample Size/100 patient	1.00	.99	1.00	.71	1.00	.99	1.00	.13

* = significant at < 0.05, ref. = reference variable, OR = Odd ratio, RTC = randomized controlled trial

Table 9. Multivariate Analysis of Survival Curve Crossings and Convergences

	95% CI for odd-ratio (OR)							
	Crossing as dependent variable				**Convergence as dependent variable**			
Independent Variables	OR	Lower	Lower	*P*	OR	Lower	Upper	*P*
Disease Type								
Cancer (ref.)	1				1			
Cardiovascular	.97	.40	2.32	.94	1.24	.39	3.93	.71
Infection	1.56	.44	5.54	.48	1.54	.30	7.74	.60
All others	1.24	.47	3.28	.65	1.94	.53	7.09	.31
Exposure Type								
Pharmaceuticals (ref.)	1				1			
Surgery	.73	.20	2.70	.64	1.15	.18	7.10	.86
Diagnostic	.44	.10	1.86	.27	.50	.10	2.35	.38
Procedure/device	.75	.32	1.78	.52	.82	.28	2.39	.73
All Others	.45	.13	1.59	.22	.90	.22	3.76	.90
Comparator Groups								
Two groups (ref.)	1				1			
Two or more groups	.46	.19	1.00	.07	.54	.17	1.64	.27
Pairwise comparison								
One pairwise (ref.)	1				1			
Two or more pairwise	1.60	.74	3.43	.22	3.70*	1.26	10.80	.02
Primary endpoint								
Death (ref.)	1				1			
Non-death Outcome	.99	.51	1.92	.98	.49	.21	1.15	.10
Secondary endpoint								
Non-death (ref.)	1				1			
Death from any cause	.71	.30	1.70	.45	8.11 *	1.00	65.48	.05
Study Types								
RCT (ref.)	1				1			
Cohort	1.19	.43	3.26	.73	.43	.12	1.45	.17
Others	.93	.29	2.88	.90	.32	.09	1.17	.08
Follow-ups in Days								
0-180 (ref.)	1				1			
181-366	.33	.10	1.36	.13	.44	.10	2.00	.29
367-732	.57	.16	1.98	.38	.65	.13	3.36	.62
733-1830	.70	.26	1.86	.47	.59	.16	2.22	.44
>1830	.76	.25	2.28	.63	.76	.18	3.23	.72
Sample size								
Sample size/ 100 patient	1.00	.99	1.00	.84	1.00	.99	1.00	.41

* = significant at < 0.05, ref. = reference variable, OR = Odd ratio, RTC = randomized controlled trial

4. Discussion and conclusion

Our research demonstrated that in major medical journals, survival curve convergences were present in almost 8 out of 10 publications and survival curve crossings in 4 out of 10 publications. Most of these convergences and crossings occurred early in the follow-up period.

Our study has a number of strengths as well as limitations. The main strength of our study lies in using the data from five major peer reviewed medical journals for one full year. The five journals are among the most prestigious medical journals and likely to present as good research as other medical journals. One of the main limitations lies in the lack of patient level data. Our study was exclusively based on published survival curves rather than original data. As a result we were unable to estimate directly the variations in relative hazards that lied behind the survival curves. We rather used survival curve convergences and crossings as an indication that the relative hazard varied over time in survival analysis. Another limitation lies in data inaccuracies because we used survival graphs rather than tables and numbers. In several publications it was difficult to ascertain whether convergence and crossings occurred or not. We observed a number of cases where two or more survival curves seemingly converge and cross each other throughout the follow-up period to the point that it was not identifiable. The problem was more complicated in survival graphs having more than two comparator groups. Another limitation lies in diversity of publications across the study characteristics. Too much variation means relatively fewer cases per category, which in turn affects the power of the statistical tests of significance. Some of the proportions in our contingency tables were overstated due to lower number of cases.

The findings of this study suggest that convergences and crossings are common phenomena in medical research. The result that was similar to the study conducted by Suciu et al in Balakrihana and Rao (eds.) 2004, where 55 (43%) out of the 127 publications they reviewed involved survival curve crossings. We found that the likelihood of survival curve convergences were twice that of survival curve crossings and most of this convergences occurred in the first six months of the follow-up period. We do not know why survival curves convergences were more common than survival curve crossings but we assume that the reason is survival curve crossing can not occur without convergences as the first step towards a full crossing. However, it has to be notice that such comparison was not considered multiple convergences and crossings that may have observed in each survival graph. We were mainly interested whether or not convergence and/or crossing involved not how many times. Our

research also revealed that convergences and crossings were hugely varied across study characteristics although we did not found any particular patterns of variations. While survival curve convergence was statistically associated with only a few explanatory variables, survival curve crossing was not associated with any of the predictor variables nether in bivariate nor in multivariate analysis. It was difficult to pinpoint exactly why this is the case. We suspect that either the numbers of cases in some categories was too small, or factors other than the study characteristics were responsible. However, we are well aware that statistical insignificance may not always mean that our explanatory variables have no association to the dependent variables.

Our study findings have at least one major implication: vertical effectiveness measures may provide results that are not representative of the effectiveness over the relevant life span for the patient. In Meta analysis researchers tends to base their result on counting adverse events at the end of trails. This practice is acceptable if the relative hazard is constant throughout the trail. Our results, however, indicates that this is frequently not the case even when we use such rough measures as survival curve crossings. Presumably, the assumption about constant relative hazards is even more frequently violated if we used more sophisticated measures. Survival curve convergences and crossings may imply that the difference in survival time is nil, grater at some times and lower at other times. If the effectiveness of an intervention to be measured only at the end of a trial, we risk losing valuable information that can save lives and improves intervention effectiveness. We, therefore, suggest that intervention effectiveness measures should be based on repeated measure, or area between the curve rather than depending exclusively on a single number generated at one specific time.

We conclude that survival curve convergences and crossings are common phenomena in medical research. The phenomena seem to be only weekly associated with particular study characteristics. The result warrants care in making inferences about the effectiveness of interventions for chronic diseases on the basis of an estimate at a single point in time.

References

Allison, Paul D. 1995. Survival Analysis Using the SAS. A Practical Guide. SAS institute Inc.

Altman DG. 1991. Practical statistics for medical research. London: Chapman & Hall.

Alwan, AD. et al. 2001. Assessment of national capacity for noncommunicable diseases prevention and control. Geneva, Switzerland, World Heath Organization (WHO).

Balakrishanan N and Rao C.R (eds.). 2004. Hand book of statistics 23. Advances in Survival Analysis. North-Holland: Elsevier.

Bland Martin and Altman G Daglas. 2004. "The logrank test." BMJ 328:1073

Catovasky, D. et al. 2007. "Assessment of fludarabine plus cyclophyllidea for patients with chronic lymphocytic leukaemia." (the LRF CLL4 trail: a randomised controlled trail).The lancet 370,230-39.

D.R. Cox and D. Oakes. 1984. Analysis of survival data. 1984. Chapman and Hall
David Hosmer and Stanely Lemeshow. Applied Survival Analysis. 1999. Wiley and Sons Ltd.

Dolan and Olsen. 2002. Distributing Health Care. Economic and ethical issues. Oxford University press

Drummond, M.F. 2005.Methods for economic evaluation of health care programmes. Oxford University Press.

Duryea, Philip J.Elias. 1992. "Tracking Survivors through the High School Years: The Theory of Survival Analysis." Paper presented at the annual meeting of the society for the scientific study of Sex (35th, San Diego, CA, November 12-15)

Gillespie and Fisher. 1979. "Confidence Bands for the Kaplan-Meier survival curve estimate." The annals of statistics. Vol. 7, No.4 920-924

WHO (http://www.who.int/topics/chronic_diseases/en/)

http://www.cancerguide.org/scurve_basic.html

Kirkwood, Betty R and Sterne Jonathan A.C. 2003. Medical Statistics. Blackwell Science Ltd.

Klein, Johan P & Moeschberger, Melvin L. 1997. Survival Analysis. Techniques for Censured and Truncated Data. Springer- Verlag New York Inc.

Kristiansen and Gyrd-Hansen. 2006. “Communicating treatment effectiveness in the context of chronic disease processes.” Expert Rev. Pharmacoeonomics outcomes Res. 6 (6), 673-679 Future Drugs Ltd.

Laakon Petter, Benestad, Haakon Breien &Olsen, Olsen, Bjørn Reino (eds.).2007. Research methodology in the medical and Biological Sciences. Elsevier Ltd.

Logan B, Klein J and Zhang M. 2008. “Comparing Treatments in the Presents of Crossing Survival Curves: An Application to Bone Marrow Transplantation.” Biometrics 64, 733-740

Motulsky, Harvey. 1995. Intuitive Biostatistics. Oxford University Press Inc.

Parmer, Mahesh K.B and Mahin, David. 1995. Survival Analysis. A practical Approch. Wiley and Sons Ltd.

Peacok, Janet and Kerry, sally. 2006. Presenting Medical Statistics from Proposal to Publications: A Step-by Step Guide. Oxford University press.

Spruance, et al. 2004. “Hazard Ratio in Clinical Trail.” American Society of Microbiology. 2004, 2787-2792 .

Therneau, Terry M and Grambsch, Patricia M. 2000. Modeling Survival data. Springer.

Wilson, Stephen A. 2004. “Interpretation of survival curves.” The Journals of Family Practice. 53:9

Yach, Derek et al. 2004. “The Global Burden of Chronic Disease: Overcoming Impediments to Prevention and Control.” JAMA 291 (21) 2616-2622.